440-WP-020-001

Earth Science Query Language Overview and Release B Design Impact Assessment

White paper - Not intended for formal review or Government approval.

May 1996

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Ron Williamson /s/ 5/2/96

Ron Williamson, Research Department Manager Date
EOSDIS Core System Project

SUBMITTED BY

Mike Daily /s/ 5/5/96

Mike Daily, Date
EOSDIS Core System Project

Hughes Information Technology Systems
Upper Marlboro, Maryland

This page intentionally left blank.

Abstract

The goal of this white paper is to document the search alternatives available to the ECS Release B design effort to define and implement a structured, language based access to ECS services. The ability to search through data collections across ECS and at a specific repository and to retrieve the results of the search is a core capability in the ECS design. The ECS requirements for search are categorized into three search domains: simple attribute, matching, and coincident.

Keywords: SQL, ESQL, Query, Threads, ECS, Search, Illustra, Release B, System Design, Data Management

This page intentionally left blank.

Contents

List of Figures

This page intentionally left blank.

1. Introduction

1.1 Purpose

This paper addresses

- the use of an Earth Science Query Language as a means for users and applications to access ECS services,
- the description of a set of queries that can be used to design the query language interfaces among the Release B subsystems, and
- the impact of the proposed ESQL baseline (Illustra SQL subset) on the Release B design, including primarily the Client, Data Management, and Science Data Server Subsystems

The constraints on the study of query languages included:

- Selection of Illustra and its associated extended SQL query language as the DBMS and baseline query language for ECS Release B
- Science user and NASA customer expectations as evidenced in the ECS Level III requirements and design working groups
- The ECS Science User Matrix

The emphasis in this version of the white paper is on the query language and the design threads that define the services, data types, parameters, sensors, sequence of operations, and data formats/types that pass between ECS clients and servers.

1.2 Review and Approval

This White Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Information Technology Systems
1616 McCormick Drive
Upper Marlboro, MD 20774-5372

This page intentionally left blank.

2. ESQL Overview

Search and access to Earth Science data has evolved from simple attribute value matching as in the V0 system prototype, through table based SQL-92 based client server interaction as in local DAAC implementations, through library catalog based search such as with Z39.50 with gateways into local databases, to ongoing attempts (by standards committees and database vendors) to integrate a full query language access to an object oriented system repository. The SQL3 and ODMG OQL standards efforts are attempting to define a query language that will provide access to complex data types and associated attributes and operations. The baseline approach chosen for ECS release B is to constrain the types of queries to those supported by Illustra's SQL extensions.

The basic ECS requirement for user access to the system is that it provide ease of use and data location transparency. The first requirement implies that the interface language (either graphical or textual) be in a form understandable to Earth Scientists and other users. Use of complex computer science terms would be unacceptable to the scientists. The use of Earth Science terms for data and algorithms that manipulate the data is therefore more desirable. Data location transparency implies that users must not have to know where data is located in order to search it. The query language must therefore be able to specify data collections independent of their location and, by implication, independent of their local schema and data format.

The ECS requirements address access to data based on geo-physical parameters (which are in some cases attributes and other cases full data types), spatial constraints and temporal constraints. The ECS design for data collections is based at the lowest level on Earth Science Data Types (ESDT) and Computer Science Data Types (CSDT). In formulating requests on the repository the assumption in this paper is that the user or application program would draw on a set of ESDT's (defined and accessible in the Data Dictionary service) with associated attributes, methods, and operations to formulate an Earth Science Query. The query could be formed in its simplest form via a graphical or forms based interface (such as the search tool in the ECS workbench) within the Client Subsystem's desktop. More complex queries (not representable via a forms or graphically) could be entered in text by users or application programs. Simple syntax sensitive editors can be used to check the validity of the constraints and the syntax of the query.

This page intentionally left blank.

3. ECS Search and Access Requirements

The overall requirements for data types, methods, and operations to be defined in the Earth Science Query Language are contained in the following documents:

- ECS Level III Requirements
- ECS Level IV Requirements for Release A and B
- ECS Data Model (Conceptual and Logical)
- ECS Release A Science Data Server PDR and CDR design documents
- ECS Release A Document Data Server PDR and CDR design documents
- ECS Release B Data Management IDR documents
- ECS Release B Science Data Server IDR documents
- ECS Release B Document Data Server IDR documents

The requirements for search services are documented specifically in the Level III requirements documents located in section 7.5.2.3 IMS Functional Requirements. As an integral part of the Release B detailed design activity each of the search specifications identified in the Level III and IV requirements must be mapped into ESQL capabilities.

The following list provides a broad overview of the ECS search requirements for distributed and site data search, for core metadata and product specific data attributes and services, and for the integration of the search results at the client's workstation. A complete list of the Level III requirements relevant to the ESQL study and a description of the impact on the ESQL are presented in Appendix A.

- Distributed and Local Search Requirements
  - Multiple Data Collections (Within a Site and Across Sites)
  - Provide Location Transparency
  - Support novice, intermediate and advanced users
- Core Metadata and Product Specific Data Values
  - Named Attributes and Services (Space, Time and Parameter)

    Geographic Metadata values and Geometric Shape Constraints

    Temporal Constraints
  - Text Field Matching (word, phrase,substring, expression)
  - Coincident Occurrences in space, time, & geo-physical parameter

- Integration of Search Results
  - Iterative Searching
  - Intermediate Heterogeneous Results

3.1 Example Queries

This section provides an example for three categories of ECS search Simple Attribute, Medium Complexity, and Coincident Search. The following section then describes in detail the related queries derived from the ECS Science Office's Modeling activities. The following example queries are included here for illustration purposes and to motivate the more detailed and comprehensive discussion in subsequent sections.

Each example describes the query in English text, provide a brief explanation of the query, and then represent the query using Structured Text with three clauses (SELECT, FROM, and WHERE), and an SQL3 example.

Simple Attribute Query (Document, Directory, Inventory)

English Text : Retrieve as hyperlinked HTML 3.0 documents the abstracts, author's name, and references of all English language research papers publish from 1990 to present that cover the topics of the correlation between sea surface temperature and the ozone hole.

Explanation: The structured query language accesses the ReferencePaper ESDT and assigns a reference X to each instance. The select clause constrains those attributes of the document that will be returned to the client as a result of matching the search constraints in the Where clause. The Select clause further specifies that each resulting document will be formatted as a hyperlinked set in HTML 3.0 format. The Where clause provides the specification of the search criteria to the document search engine.

Structured Query Language Text:

```
SELECT  (X.author, X.abstract, X.references).hyperlink("HTML 3.0")
FROM  ReferencePaper X
WHERE  X.PublicationDate > 1990 AND X.language = {English}
     AND X.Topics = {correlation sea surface temperature ozone hole}
```

SQL3:

Schema:

```
CREATE TABLE ReferencePaper
 (date DATE, author STRING, abstract VARCHAR(4096),
  language STRING, topics SET(STRING))
```

Query:

```
SELECT  author, abstract, references
```

FROM ReferencePaper

WHERE date >= "1990/01/01"

AND language = "English"

AND CONTAINS(topics, "correlation")

AND CONTAINS(topics, "sea surface temperature")

AND CONTAINS(topics, "ozone hole"))

Note:

STRING is a built-in ADT for SQL3, so is SET type constructor. CONTAINS is a function defined for the SET type.

Medium Complexity Query

English Text: Select all stations from the CALNEVA dataset with precipitation > 6 inches; sort by date; Use the dates to look at data in Frank Gehrke's SNOW database and to derive for each date snow ratio=snow depth/precipitation. Explanation: This query involves the joining of two separate dataset, the CALNEVE precipitation dataset and the Gehrke SNOW dataset. The joined values of snow depth and precipitation on the same date at the same station are used to calculate the snow ratio at each station on each separate data. The resulting list is then sorted by station and by date.

Structured Query Language Text:

SELECT (X.date, SnowRatio = (X.snowDepth / Y.precipitation).sortBy(X.date)

FROM SNOW X, CALNEVA Y

WHERE X.date = Y.date

AND Y.precipitation.asInches() > 6

SQL3:

Schema:

CREATE TABLE SNOW (date DATETIME, snowDepth FLOAT, station ST_POINT);

CREATE TABLE CALNEVA (date DATETIME, precipitation FLOAT, station ST_POINT);

Query:

```
SELECT X.date, X.snowDepth/Y.precipitation
  FROM SNOW X,  CALNEVA Y
 WHERE X.date = Y.date
      AND X.station = Y.station
      AND Y.precipitation > 6.0
ORDER BY X.date;
```

Note:

This assumes the snowDepth and precipitation are already measured in inches. If it's not the case, we can have a new ADT for the measurements, and conversion functions like INCH or METER cane be defined. ST_POINT is a spatial ADT being defined in the SQL/MM package.

Coincidence Search Query

In this query we try to demonstrate the kinds of complex coincidence searching that need to be specified by Earth scientists while conducting their inter- disciplinary research.

English Text: Find all examples of ocean blooms in Sea-WiFS ocean color data which are coincident with ozone concentrations of less than 50% for the period starting in 1990 to the present and at southern latitudes south of 55 degrees south.

Explanation: There is some evidence that toxic ocean blooms are linked to ozone depletion. In this query the GOMR data on ERS-2 will be searched to identify areas of ozone depletion (i.e the so-called ozone hole) below the defined value (this will vary temporally). Then use this dynamic location data to direct the use of a "bloom" detection algorithm in the Sea-WiFS data which is coincident with the ozone depletion. Note that this assumes a content based search (either algorithmically or via manual inspection) has been performed on the data either at the client's site or on ingest to determine the ozone concentration contours in the GOMR data and similarly for the identification of the spatial extent of ocean blooms in the Sea-WIFS ocean color data. Content based searching by the ECS is not part of the ECS baseline. ECS will provide product data to requesting clients and any content based search and subsetting will be conducted at the client's site.

SQL3:

Schema:

```
CREATE VALUE TYPE t_SEAWIFS
( PUBLIC name  STRING;
  timeStamp DATETIME;
  data  ST_SPATIAL_RASTER;
```

```
      PUBLIC FUNCTION OceanBlooms(seawifs t_SEAWIFS) RETURNS
SET<ST_SPATIAL_RASTER>
  BEGIN
  ... (code for finding ocean blooms and return
   the result as a set of small raster images)
 END
);
CREATE VALUE TYPE t_GOMR
( PUBLIC name  STRING;
 timeStamp DATETIME;
 data  ST_SPATIAL_RASTER;

 PUBLIC FUNCTION OzoneSelect(gomr t_GOMR, concentraion_percent integer)
RETURNS SET<ST_SPATIAL_RASTER>
  BEGIN
  ... (code for identifying regions that has ozone
   concentration level above specified
   percent, and return the result as a set
   of small raster image)
 END
);
CREATE TABLE GOMR (data_set t_GOMR);
CREATE TABLE SEAWIFS (data_set t_SEA_WIFS);

Query:
SELECT  OceanBlooms(X.data_set)
FROM  SEAWIFS X, GOMR Y
WHERE SpatialOutline( OceanBlooms(X.data_set) ).South < -55.0
     AND X.data_set.timeStamp >= "1990/01/01"
     AND Y.data_set.timeStamp = X.timeStamp
     AND SpatialOutline( OceanBlooms(X.data_set) ) overlaps
 SpatialOutline( OzoneSelect(Y.data_set, 50) )
```

Note:

ST_SPATIAL_RASTER is a SQL3 ADT which supports a raster image (2-d array of binary data) with spatial information. SpatialOutline is one of its functions. Here 2 new ADTs are created for SEA WIFS and GOMR data set to implement their user-supplied functions: OzoneSelect() and OceanBlooms().

4. Query Thread Definitions

Each query thread is described in English with specific references to the Data Model's Pyramid Layers, the specific Earth Science Data Type (ESDT) references (if any), and the geo-physical parameter and instrument name. An SQL3 query description is supplied for those thread steps that include a parameterized search of the ECS data collections.

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread 1			
User requests (via telephone) all general documentation on the TOMS instrument and technical documents characterizing the bandpass (WEB Keyword SEARCH) or Select X.generalDoc, X.techDoc From BibRef X Where X.instrument = TOMS AND X.techDoc(keyword) = "bandpass"	Bibliographic References(ECS & Non-ECS)	NA	NA TOMS
Thread 2			
User phones and wants to place an order. USP performs search. Arizona - South East of Phoenix on date = July 28, 1998, 24 hours centered on 0500 UT. Parameters are flash data, observation time , single events Select X. flash_data, X.obs_time From LIS02 X, LIS03 Y Where SpatialIntersect(Map(Arizona), Map(Phoenix).southEast(), Set(X.location, Y.location)) AND X.obs_time.centered_on (24 hrs, "1998/07/28 : 0500 UT" AND Y.obs_time = X.obs_time AND X.type = Y.type = "single event"	Inventory	LIS02, LIS03	Flash data, Observation time, Single events LIS
Thread 3			
User Services Person asks student about specific information he desires. Precipitation, Leaf-Area-Index, Regional Global Change Susceptibility Index for a minimum 5-year continuous data. Select * From Precipitation X, LAI Y, RGCSI Z Where X.continuous_sample(5, "yr") AND Y.continuous_sample(5 , "yr") AND Z.continuous_sample(5, "yr")	Inventory(ECS & Non-ECS)	NA	LAI, PRECIP., NDVI, Global Change Susceptibility Index

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread 4			
User requests info. on data available related to cross reference between forest modelling algorithms and data produced via terrestrial measurement techniques. (WEB Keyword SEARCH) or Select X.document From Guide X, Guide Y Where CrossReference(X.keywords("forest modeling algorithms), Y.keywords("data produced terestrial measurement"))	Guide(ECS & Non-ECS)		Forest model, terrestrial measurements of CO2, biomass, temperature
Thread 5			
Locate electronic journal (WEB Keyword SEARCH) or Select X.document From Journal X Where X.keywords("reference1", ...)	Biblio-graphic References		
Thread 6			
User logs on to ECS and specifies latitude and longitude of his study area. He requests a search for info. on the following types of data for the specified geographic area. Land Cover classes, Land surface reflectance values, Vegetation Indices data. Select * From Directory X Where X.spatial_outline(lat, long) AND X.parameter = "land cover classes" AND X.parameter = "land surface reflectance" AND X.parameter = "vegetation indices"	Directory		landcover classes, land surface reflectance, vegetation indices

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread 7			
User generates a query for highest resolution Landsat Thematic Mapper (TM) data available for a 512 x 512 km region centered at 42 deg N latitude, 112.5 W longitude. Select X.data From Landsat-TM X Where X.spatialIntersect(Rectangle(42N, 112.5W, center, 512 km NS, 512 km EW) AND Max(X.resolution)	Inventory	Landsat - TM	? Enhanced Thematic Mapper
Locate 1 Km. AVHRR level 1A data; Same geographic region as in Step 1; Temporal range = (20 days before the day of the chosen Landsat Image plus 60 days following the Landsat image) = 80 days Select X.data From AVHRR_1B X, Landsat-TM Y Where From Where X.spatialIntersect(Rectangle(42N, 112.5W, center, 512 km NS, 512 km EW) AND Y.spatialIntersect(Rectangle(42N, 112.5W, center, 512 km NS, 512 km EW) AND Max(X.resolution) AND X.time_stamp >= Y.time_stamp - 20 AND X.time_stamp <= Y.time_stamp + 60	Inventory	AVHRR -Level 1B	AVHRR

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread 8			
Search for the QA statistics describing the validation of the products. Select X.QA, Y.QA From AST02 X , AST03 Y Where X.validation = true AND Y.validation =true	QA Stats	AST02, AST 03	NA ASTER
User places a standing order for 18 months worth of MODIS data to be delivered once a month via 8mm tape Select Monthly (X.data(8mm), Y.data(8mm), Z.data(8mm)) From MOD13 X , MOD14 Y, MOD17 Z Where X.validation = true AND Y.validation =true ANDSet (X, Y, Z).time_range(present, 18 months)	Level 2	MOD13, MOD14, MOD17	MODIS
Thread 9			
Locate summary statistical info for several datasets that deal with NDVI as a measure of vegetation productivity for the same area previously mentioned. Select * From Summary X Where X.parameter= NDVI AND X.spatialOutline(area_of_interest)	Summary Stats (ECS & Non-ECS)	NA	NDVI
Thread 10A			
User queries ECS about availability of the following data for the Little Washita watershed, OK - Soil maps, DEM, Insitu rainfall data, Run off data Select * From Catalog X Where X.spatialIntersect (Map ("Little Washita Watershed") AND X.parameter = { Soil maps, DEM, Insitu rainfall data, Run off data }	Inventory	NA	Soil, DEM, Insitu rainfall data

Description	Pyramid Reference	ESDT Reference	Parameter Instrument Name
Thread 11 A and 11B			
User generates queries on inventory data about Level 1B SAR images in 100 x 100 km squares for the location identified in the previous step, for three specific time ranges that each span approx. 3 days. Select * From Inventory X Where X.product = "SAR" AND X.level = "1B" AND X.format = 100kmSquares AND X.spatialOutline(area_of_interest)	Inventory	SAR (level 1b)	NA SAR
User generates searches on Guide data to get information on SSM/I, gridded brightness, and temperature data sets. (WEB Keyword SEARCH)	Guide	NA	Brightness, Temperature
Thread 12			
User requests a search of all Directory information on all Level 1 data over Greenland for 1 month. Select * From Directory X Where X.level = "1" AND X.spatialOutline(Map("Greenland"))	Directory	NA	
Thread 13			
User searches Algorithm information for Algorithm Theoretical Basis Documents (ATBD) for CERES products. (WEB Keyword SEARCH)	Algorithms	NA	NA CERES
Thread 14			
Users generates a search query for all Landsat Thematic Mapper scenes for the Sierra Nevada during winter. Select * From Catalog X Where X.product = "Landsat TM" AND X.spatialOutline(Map("Sierra Nevada") AND X.timeperiod(Calendar(Winter, All))	Inventory	Thematic Mapper	NA Landsat

Description	Pyramid Reference	ESDT Reference	Parameter Instrument Name
Thread 15			
User enters "LIS" and "Continental U.S. and Coastal Areas" as search criteria for GUIDE information on LIS parameters. (WEB Keyword SEARCH)	Guide	NA	NA LIS
Thread 18			
User Queries ECS for availability of following types of data for his subbasin of interest : 1)DEM 2)Land Cover 3) Vegetation Indices 4) Soil Moisture 5) Snow Cover 6) Snow Water Content. Select * From Inventory X Where X.product = Set(DEM, LandCover, Vegetation Indices, Soil Moisture, Snow Cover, Snow Water Content)	Inventory	NA	DEM, Land cover, Vegetation Indices, Soil moisture, Snow cover, Snow depth, Snow water content
Thread 20			
User requests a search for the following types of data: snow cover, snow thickness, vegetative indices, sea ice, cloud properties, precipitable water, rainfall, surface radiative fluxes	Level 2	NA	snow cover, snow thickness, vegetative indices, sea ice, cloud properties, precipitable water, rainfall, surface radiative fluxes ASTER, CERES, MODIS
Thread 22A and 22B			
User logs on and requests a search for Level 0 MODIS data for channels 7,20,29, and 31, for the last 3 satellite passes over the northern 3rd of the Phillipines.	Level 0	MOD00	NA MODIS

Description	Pyramid Reference	ESDT Reference	Parameter Instrument Name
Thread 23A			
User searches Inventory for Level 4 DAO 4D assimilation products for latitudes above 35 deg N and latitudes below 35 deg S, AND NMC station meteo. data for yymmdd in North America, and ECMWF forecast for yymmdd regions specified by lat/long points. Select X.dataset, From DAOAssim_L4 X, NMC Y Where (X.location >= 35N OR X.location <= 35S) AND Y.parameter = meteorological AND Y.date = yymmdd AND Y.location.intersects(Map(North America)) AND Z.date = yymmdd AND Z.location.intersects(Region(lat/long list))	Inventory	NA	NA

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread 24			
User draws a polygon on the map (provided by IMS interface) defined by the following lat long: 10'S - 40'S, 70'W - 90'W, and requests system to search for all Level 1 radiance data for this area with: <= 1km res. and a 1 yr. temporal coverage. Select * From Catalog X Where X.location.intersects(Region(10S, 40S, 70W, 90W) AND X.level = 1 AND X.parameter = radiance AND X.resolution <= 1km AND Cumulative_Coverage (X.date, 1yr)	Inventory	NA	Radiance

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread: Coincident Search			
User queries system for TOPEX Sea Surface Variability, MODIS Ocean Color, and AVHRR Sea Surface Temperature data for the "rectangular" geographic area described by the following coordinates: (degrees.minutes) 34.20N,122W to 34.20N,121W, and 34N,122W to 34N,121W for April 20, 2001 to April 30, 2001 Select * From Catalog W, TOPEX X, MODIS Y, AVHRR Z Where X.parameter = "sea surface variability" AND Y.parameter = "ocean color" AND Z.parameter = "SST" AND W.spatialIntersect(X.spatial_outline, Y.spatial_outline, Z.spatial_outline, Rectangle (coords)) AND temporalIntersect(X.time, Y.time, Z.time, TimeRange("4/20/2001", "4/30/2001"))	Inventory	MODOCCLR_L3_DY, Multi-Channel Sea Surface Temperature (MCSST), TOPEX Sea Surface Variability	Sea Surface Variability, Sea Surface Temperature, Ocean Color AVHRR, TOPEX, MODIS
User queries system for Digital Elevation Model (DEM), Soil Moisture, Precipitation, Soil type, Snow Cover, and In-Situ stream level (weir) data for the area 76.96W to 76.3W and 40.75N to 40.58N for January 1, 2000 to January 24, 2000 Select * From DEM W, Catalog X Where X.parameters = { soil moisture, precipitation, soil type, snow cover, in-Situ stream levels) AND Set(W, X).spatialIntersection (Rectangle(76.96W, 76.3W, 40.75N, 40.58N)) AND Set(W,X).timePeriodIntersection(1/1/2000, 1/24/2000)	Inventory	AST_14, MOD10_L2, STATSGO, MOD05_L2	Soil Moisture, Precipitation, Soil Type, Snow Cover, Stream Level ASTER, MODIS

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread: Coincident Search			
User queries system for Atmospheric Data for the Month of July for 4 yrs, 1999-2002, two areas: (degrees) Area 1) 30N 100W to 30N 105W and 25N 100W to 25N 105W; Area 2) 30N 105W to 35N 105W and 35N 110W to 30N 110W Parameters of interest: precipitable water, aerosols, ocean surface characteristics Select * From Catalog X Where X.discipline = atmospheric AND X.parameters = {precipitable water, aerosols, ocean surface characteristics) AND X.periodicTimeIntersection(Month, July, Duration, 1999, 2002) AND (Set(W, X).spatialIntersection (Rectangle(30N, 100W, 30N, 105W, 25N, 100W, 25N, 105W)) OR Set(W, X).spatialIntersection (Rectangle(30N, 105W, 35N, 105W, 35N, 110W, 30N, 110W)))	Inventory	MIS_2AS, MIS_3AM, MIS_3SM, MOD04_L2, MOD38_L2	Aerosols, Surface Parameters, Precipitable Water MISR, MODIS
User queries the system for data for the time period February 1, 1999 to September 30, 1999 for the area within the following coordinates: 76W36N 75W36N, 77W35N 76W34N, 79W34.5N 78W33N, 81W33N 79W32N, 83W30.5N 80W30.5N. Parameters of Interest: sea surface temperature, Normalized Difference Vegetation Index (NDVI), Leaf Area Index (LAI), Fractional Photosynthetically Active Radiation (FPAR), Precipitable Water, land cover, ocean color, soil type SQL3 is similar to that defined above.	Inventory	MOD28_L2, MOD09_13_L2, MOD15_L4_10DY, MOD05_L2, MOD12_L3, _96DY, MODOCCLR_L3_COMP_WK, SSM/I, STATSGO	sea surface temperature, NDVI, MVI, LAI,FPAR, Precipitable water, Land cover, Ocean color, soil type MODIS, SSM/I, CERES

Description	Pyramid Reference	ESDT Reference	Parameter Name Instrument Name
Thread: Coincident Search			
User queries system for data for the time period March 1 to September 30, 2001 for the locations: 77W to 76 W and 41N to 40.6N and 80W to 73W and 35.5N to 44N. Parameters of Interest: topography, precipitable water, land surface temperature and emmissivity, Normalized Difference Vegetation Index (NDVI), soil data, monthly TOA and surface radiation, precipitation SQL3 is similar to that defined above.	Inventory	MOD05_L2, MOD11_L2, MOD09_13_L2, AST_05, AST_08, CER06bAP, STATSGO	Topography,Total precipitable water, Land surface temperature & emmissivity, Normalized vegetation difference index, Normalized vegetation difference index, STATSGO (State Soil Geographic Database) data, Monthly TOA and surface radiation, Precipitation MODIS,ASTER, CERES,SSM/I, AVHRR

This page intentionally left blank.

5. Conclusions and Summary

The Release B baseline design for the development of a structured language representation of a user search specification (i.e. an ESQL based query) is to use a subset of the Illustra SQL extension in support of data type extended relational models. The longer range plan is to target the SQL3 standard as it evolves and matures but to focus on and influence the Illustra implementation of a subset of the SQL3 specification. The Release B baseline for querying includes basic attribute, spatial, and temporal expressions. Full coincident search capacities will be included in later releases.

Both the SQL3 standard and the Illustra SQL extended with data types provides for a wide range of boolean expressions including attribute value matching across data types and integration of data type operations with the query expression. The Release B design and prototyping teams will map the search requirements onto the capabilities provided by the Illustra query language. The resulting subset will then be documented as part of the prototyping and detailed design, implementation, integration, and test phases of Release B.

Three Release B subsystems are primarily impacted by the integration of a structured Earth Science Query Language: Client, Data Management, and Data Server. The design issues discussed in this section include:

- Client
    - Look and Feel of the User Interface to include more complex boolean expressions, matching queries, and coincident search specifications
    - Advanced user support of complex query specifications not expressible using the graphical user interface
    - Interface between the Client and the data management and data server API's (specifically the Data Server Interface and the Query Object Defintion)
- Data Management
    - Parser for the ESQL based Query Object.
    - Mapping of Conceptual Schema onto the Data Server external schemas
- Science Data Server
    - Query Object Management via the Distributed Object Framework and/or the Illustra DBMS Type Manager
    - Mapping of Conceptual Schema onto the Internal Schema within Illustra

The design issues in each of these subsystems related to query management are in large part independent of the use of an ESQL or a GLParameterList representation of the user query. The design impact centers on the creation, parsing, and interpretation of the Query Object. The Query Object should be design to encapsulate the query specification, to provide interfaces to

create standard query clauses such as in the SQL Select, From, and Where clauses, and to support sub-query creation by the DIM and LIM.

The Client Subsystem is responsible for transforming user inputs (either via the GUI or via text based input) into an instance of a query object that contains i) the specification of the results to be returned (e.g. as specified in the Select clause), ii) the specification of the data types from the conceptual model (e.g. ESDT's), and iii) the boolean expression relating attributes, values, and relationships among data types. The Release B requirements drive the content of the query expression, such as the complexity of the boolean expression, the inclusion of matching type queries, and the exclusion of coincident search. The use of ESQL as the internal representation is independent of the content of the required queries. A common query representation within the Client subsystem would be desirable but not critical. The change in representation to an ESQL syntax could occur at the interface to the search servers (such as the DIM, LIM, or Science Data Server). If a syntax directed editor approach is chosen, as illustrated in Figure 1, the syntax of the chosen ESQL (Illustra SQL in Release B) must be embedded in the GUI interface drivers to provide a user friendly interface to the syntax of the query language.

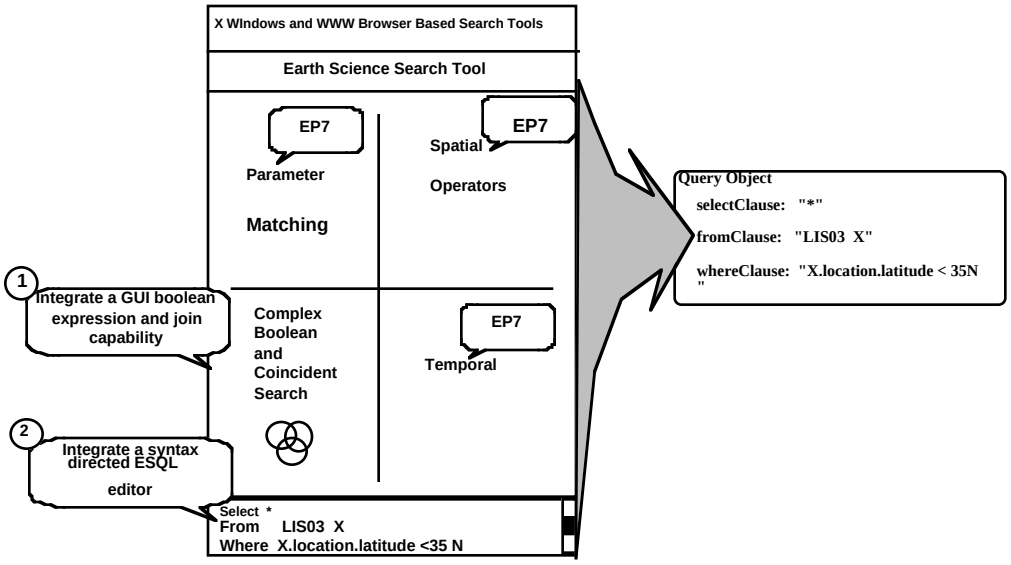

Figure 5-1. GUI Client Extensions in support of ESQL

The Data Management subsystem is responsible for the maintenance of the data definitions via the data dictionary service, the federation of schema and parsing of system wide queries within the DIM, and the integration of schema and parsing of site specific queries within the LIM. The data dictionary service supports the user via the Client workbench in formulating syntactically valid queries and queries with valid attribute values and dependencies. The conceptual model as viewed by the user is maintained by the data dictionary service and thus feeds the inputs to the

Select, From, and Where clauses of the query expression. Both the DIM and the LIM must parse the content of the incoming Query Object and therefore must understand the syntax of the query expression and the data types associated with the results. In addition, any sub queries created as a result of parsing the incoming query must be generated using ESQL syntax and then passed to the relevant search servers.

The Data Server subsystem is composed of the Science Data Server and the Document Data Server. The Document Data Server is designed primarily as a World Wide Web server with text based keyword and proximity search servers. Documents can be searched using simple data types for the documents, and simple keyword based expressions. The primary impact of the choice of a query language is on the Science Data Server, and specifically on the Distributed Object Framework layers in the server. The underlying database management system for access to the data types contained in the "higher" levels of the data pyramid is baseline in Release B as Illustra. Assuming a syntactically correct query expression is passed from the Client Subsystem to the Data Server Subsystem, the query expression may be passed directly through to the Illustra server with minimal impact on the Science Data Server design. The primary trade-off to be made is the balance between Illustra Data Type design and implementation and the Distributed Object Framework request management objects.

A prototype design of the query language is presented next. The design description is presented as a top level query, followed by and object Model identifying the key objects and their interfaces. The query threads overlay on top of the object model as an object interaction and event trace diagram showing the dynamic flow through the prototype design.

The query thread example for the prototype design is taken from the coincidence search examples in the tables above:

User queries system for TOPEX Sea Surface Variability, MODIS Ocean Color, and AVHRR Sea Surface Temperature data for the "rectangular" geographic area described by the following coordinates: (degrees.minutes) 34.20N,122W to 34.20N,121W, and 34N,122W to 34N,121W for April 20, 2001 to April 30, 2001

```
Select *
From  Catalog W, TOPEX X,  MODIS Y,  AVHRR Z
Where
   X.parameter = "sea surface variability"
AND  Y.parameter = "ocean color"
AND Z.parameter = "SST"
AND
  W.spatialIntersect(X.spatial_outline,
       Y.spatial_outline,
```

```
        Z.spatial_outline, Rectangle (coords))
    AND
     temporalIntersect( X.time, Y.time, Z.time,
                    TimeRange("4/20/2001", "4/30/2001") )
```

The first element of the prototype design is the key object diagram for the three subsystems of interest: Client, Data Management, and Science Data Server, reference Figure 2. The key object associated with the ESQL design is the Query Object that includes a specification of the query specification, for example using and IDL specification:

```
interface  Query {
        selectClause( DDString selectString);
        fromClause(DDString fromString);
        whereClause(DDList booleanExpression);
};
```

In the definition of the Query object the DDString and the DDList types are derived from the term definition defined in the Data Dictionary. As the Client's ESST object collection build the query strings, the data dictionary term definitions, valid values, and dependent valids are used to ensure proper term usage. These terms are tagged by the data dictionary and encapsulated with the Query object. The encapsulation of data dictionary term information simplifies the query parsing within the Data Management and Science Data Server subsystems. The Advertising Service is indirectly impacted by the ESQL in that any of the servers supporting an ESQL interface and associated protocols will advertise their ESQL capabilities via the advertising service.

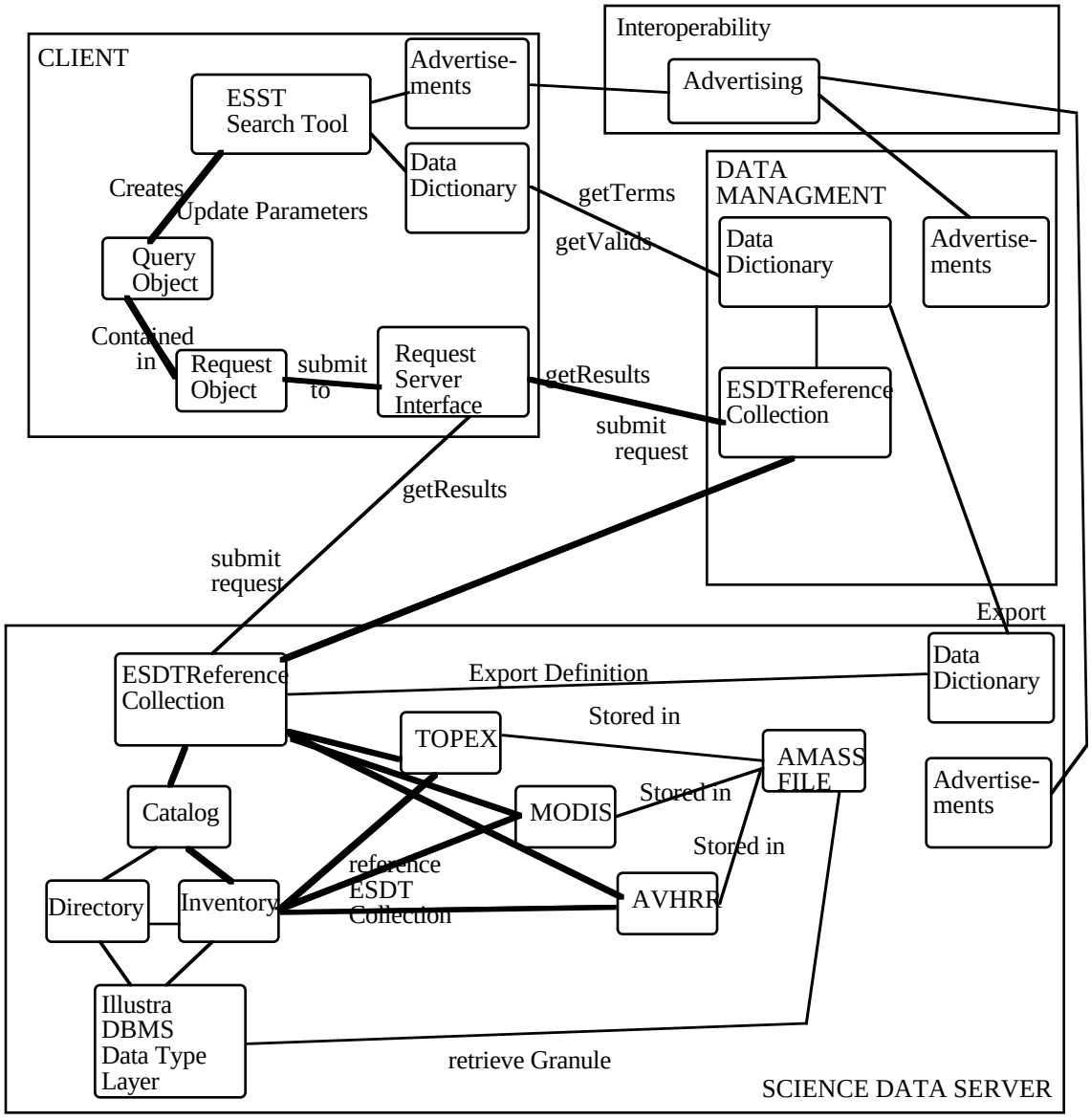

Figure 5-2. Prototype Design Object Model

This page intentionally left blank.

6. Suggested Reading

ECS, 1994: Summary of the ECS System Design Specification. http://edhs1.gsfc.nasa.gov.

Elkington, M., Meyer, R. and McConaughy, G., 1994: Defining the Architecture of EOSDIS to Facilitate Extension to a Wider Data Information System. ISPRS'94, Ottawa.

Enloe, Yonsook, August, 1994, ECS Technical Notes, Z39.50 and SQL3 Evaluation.

Fox, S., 1994, EOSDIS Core System Science Information Architecture, FB9401V2. URL: http://edhs1.gsfc.nasa.gov

Heller, Denise, March, 1994, ECS Technical Notes, Query Classes.

Moxon, B. 1993: 19300611 Science-based System Architecture Drivers for the ECS Project. URL: http://edhs1.gsfc.nasa.gov.

Tyahla, L., 1994, ECS User Characterization Methodology and Results, 194-00313TPW. URL: http://edhs1.gsfc.nasa.gov

West, S., 1994, ECS Scientist User Survey, 19400549TPW, URL: http://edhs1.gsfc.nasa.gov

Wingo, T., 1994, User Characterization and Requirements Analysis, 19400312, URL: http://edhs1.gsfc.nasa.gov

This page intentionally left blank.

A. Requirements Summary and ESQL Impact

L3 Req.	Requirement Description	Impact on ESQL
IMS-0260	The IMS shall provide interactive and batch information management capabilities for authorized users to add, update, delete, and retrieve information from the IMS data bases.	Retrieval & Update
IMS-0272	The IMS shall provide the capability for users to search the information on the science processing library holdings.	Retrieve SPL ESDT
IMS-0273	The IMS shall provide the capability for users to order science processing library holdings.	Order SPL ESDT
IMS-0290	IMS internal data base management queries shall be expressed in a standard query language.	Standard QL
IMS-0320	Standard Product related metadata shall contain, at a minimum: Keywords and glossary from investigators Keywords, synonyms, and glossary for cross-product and cross-directory referencing Identifiers for locating products in the DADS archive by granule Documentation on algorithms, including version history, authors, written description of product, equations, and references Documentation on instrument(s) and spacecraft(s) including history of housekeeping and ancillary parameters, discipline characterization, calibration parameters, key individuals, and references Identifiers, algorithms, written descriptions, equations, authors, and references associated with static browse products and subsetted, subsampled, and summary data products Published papers, research results, "significant" results, and references by author and date Key organizations and personnel for all product-related DAACs, ADCs, and ODCs Granule-specific information as listed in Tables C-10 and C-11 in Appendix C	Metadata ESDTs Supported in ESQL References Keyword Attributes Documents Algorithm Instrument Spacecraft Papers Personnel Granule Attributes
IMS-0330	The metadata maintained by the IMS shall provide a cross reference that relates science data to the following at a minimum: Calibration data, navigation data, and instrument engineering data Processing algorithms used for data generation at the PGS Software used for data generation at the PGS Parameters used for data generation at the PGS Input data used for data generation at the PGS Data recipients The PGS at which the data was processed QA and validation data, reports, and algorithms	ESDTs Supported in ESQL Calibration Algorithms Software Parameters Input Data Recipients Processor QA Validation
IMS-0340	The metadata maintained by the IMS shall contain content-based summary information, including statistical summaries and granule features, for all ECS standard and special products.	ESDTs Supported in ESQL Summary

L3 Req.	Requirement Description	Impact on ESQL
IMS-0350	The IMS shall provide the capability for authorized personnel to add, delete, or modify ECS metadata entries, individually or in groups.	
IMS-0355	The metadata shall be expandable to include additional attributes which are identified during the mission and deemed useful for data search.	ESQL Attribute access updatable
IMS-0500	The IMS shall provide access to information to include at a minimum: Metadata, Spacecraft housekeeping and ancillary data, engineering data, EOC historical data, data acquistion plans and schedules, processing schedules, documentation, ESDIS project policies and procedures obtained from SMC db, science processing library software, documentation on dat format and metadata standards	ESQL access to Data Pyramid Range of ESDTs
IMS-0530	The IMS shall provide document text search	Document ESDT Access
IMS-0540	The IMS shall display PGS system processing schedules to users.	System Processing Schedule ESDT
IMS-0550	The IMS shall allow a user to locate and identify desired data without detailed knowledge of the ECS's: Architecture, Database management system, database structure, query languages, data formats	Declarative Interface GUI I/F
IMS-0560	The IMS shall decompose complex data base search requests into executable data base queries in a manner which is transparent to the user.	ESQL Query Processor
IMS-0570	The IMS shall provide an incremental search capability.	Incremental ESQL Scripting
IMS-0580	The IMS shall provide geographic overlays to aid in the selection of spatial data and to enhance the display of metadata.	ESQL link to Graphics Interface
IMS-0600	The IMS shall provide the capability to search the Global Change Master Directory of information that describes whole EOSDIS, non-EOSDIS, and ADC earth science data sets.	ESQL access to Master Directory
IMS-0610	The IMS shall provide the capability to search the data inventory which describes each granule of EOSDIS data.	ESQL access to Inventory View
IMS-0620	The IMS shall provide access to inventories of ODCs and ADCs selected from paragraph 4.3.4 via level II and level III catalog interoperability as specified in ICDs.	ESQL acces to ODC and ADC
IMS-0630	The IMS shall provide the capability to select metadata for retrieval by: Boolean operators, Relational operators, attribute values, combinations	ESQL support for Boolean combinations
IMS-0640	The IMS shall provide the capability to query geographic metadata by any of the following criteria at a minimum: Geographic reference, Data element content (as specified in metadata), minimum bounding rectangle, point and radius, Geographic name (based on a standard data base, such as USGS Geographic Names Information System)	ESQL specification of geo query
IMS-0650	The IMS shall query non-geographic metadata by any of the following criteria at a minimum: Exact word match, Phrase match, Character set (string), Wildcard construct (prefix, embedded, suffix), Character range, Logical and Boolean operators, Min/max range search, any combination.	ESQL specification of word match, and string search

L3 Req.	Requirement Description	Impact on ESQL
IMS-0660	The IMS shall provide inventory metadata search based on any combination of the core inventory (Table C-10, Appendix C) metadata attributes and geophysical parameters at a minimum.	ESQL specification of Core Inventory Search
IMS-0666	The IMS shall provide informational messages to indicate that a query is being executed.	ESQL Search Status
IMS-0667	The IMS shall provide the capability for the user to terminate a time-intensive query before all hits are found, or to abort any time-intensive operations.	ESQL Search Control
IMS-0670	The IMS shall provide the capability to accept, validate, and fill orders from users for periodic delivery of information stored at the IMS.	ESQL Access to Retrieve and Order Operations
IMS-0700	The IMS shall provide the capability for users to browse subsetted, subsampled, and summary data products, which have been processed at the PGS during the routine production processing and archived at the DADS, whenever associated inventory information is displayed.	ESQL Access to Browse Operations & Products
IMS-0720	The IMS shall provide the capability to order subsetted, subsampled and summary data products which are processed ~~ad hoc~~ using pre-existing processes in response to user requests for: Temporal subsetting, Spectral subsetting, Spatial subsetting with rectangular boundaries, Pixel subsampling	ESQL Access to Subset, Subsampled Summary Operations & Products
IMS-0810	The IMS shall prepare, for output to the DADS, product orders to retrieve specified data from the archive and distribute it, which contains the following information at a minimum: Requestor id, data type, data set identifier, dataset subsetting instructions, data formats, distribution instructions, including media requirements, request priority, suggested earliest start time, suggested latest completion time.	ESQL Access to Product Attributes for Order/Retrieval
IMS-0820	The IMS shall provide to the user product order status information from the DADS to confirm or reject an order, which contains the following information at a minimum: Requestor identification, request identification, start status	ESQL Status Display
IMS-0920	The IMS shall provide the capability for users to construct and submit standing orders and one-time requests for processing of ECS data by pre-existing processes, which shall contain the following information at a minimum: Requestor identification, Algorithm input requirements, text description of need for processing, Level 0-4 dataset subset, required time of generation, requested priority for product processing, resulting product type, distribution instructions (shipping information, media requirements).	ESQL Specification of Standing Orders
IMS-0930	The IMS shall provide the capability to search metadata holdings for the purpose of identifying the product desired and the input data to be processed.	ESQL Specification of all holding
IMS-0940	The IMS shall integrate the searching of metadata holdings for identifying information needed to complete a processing request into the request construction and submission process.	ESQL Specification Integrated into Processing Request

This page intentionally left blank.